DOCUMENT RESUME

ED 209 995 HE 014 563

AUTHOR Zimmerman, M. Scott; Doolittle, Allen E.
TITLE First Semester Retention of University of Iowa Students.
PUB DATE [81]
NOTE 18p.
EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Academic Aptitude; *Academic Persistence; Aptitude Tests; *College Freshmen; Educational Background; Evaluation Criteria; Factor Analysis; Higher Education; Institutional Research; Predictive Measurement; *Predictor Variables; Research Design; Student Attrition; *Student Characteristics
IDENTIFIERS *University of Iowa

ABSTRACT

Pre-enrollment characteristics that distinguish nonpersisting students from students who persist into the second semester of their freshman year were studied at the University of Iowa in 1978. Three sources of information were used to obtain predictor variables: American College Testing (ACT) program test scores, the ACT Student Profile, and high school transcript information from the registrar. Study objectives were to reduce the large number of pre-enrollment variables to a small number of useful prediction variables and to use linear and nonlinear classification techniques to exhaust the predictive power of the predictor variables. Of the 2,850 entering freshmen in the fall 1978 cohort, there was enough information to allow factor scoring for 1,711 of the students. It was found that the difference between using 43 factors and 23 factors was negligible. There were no indications from this study or previous studies that a larger number of the pre-enrollment variables could account for more than variance of persistence; however, the 23 factors did not predict retention any better than an earlier study that used only three variables. It is concluded that pre-enrollment variables per se did not directly predict persistence. It is recommended that the smallest set of factors that maintains the current level of prediction of persistence should be developed and used in retention studies to save on costs of data storage and analysis. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209995

First Semester Retention of University of Iowa Students

H. Scott Zimmerman and Allen E. Doolittle
Evaluation and Examination Service
The University of Iowa

1/8

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Allen E. Doolittle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 014 563

INTRODUCTION

To increase their probability of survival, colleges and universities need to increase student enrollment. Student enrollment is increased by reducing the proportion of students who withdraw or by increasing the number of incoming students. The retention of current students is more cost effective than raising the number of admissions. Summerskill (1962) and Pantages and Creedon (1978) have shown that 50% of an institution's entering students do not graduate from that institution. Forty percent of this non-persisting group transfer and eventually graduate from another institution. The remaining 60% of the non-persisters never graduate. The departure of such a large number of students is very costly to the institution in terms of lost revenues and to many of the students in terms of satisfaction with their educational experience. If potential non-persisting students could be accurately identified, institutions would have the opportunity to target these students for additional help (advising, counseling, adjustment programs, etc.) and thereby increase student retention.

The term retention has been defined in terms of the period of time used to classify persisters and non-persisters. Definitions of retention for classification of per-

sisters and non-persisters have ranged from continued enrollment over one semester for first semester freshmen to enrollment for a period of 4 and 5 years or graduation. Retention research at The University of Iowa has encompassed the entire range of definitions. Unfortunately, none of these studies were able to accurately determine the variables predicting retention or accurately classify students as persisters or non-persisters. The lack of success of these studies may have been due to the heavy use of pre-enrollment data such as ACT scores and High School transcript information. This study focuses only on retention in the first semester of college. This choice of the time period would presumably allow the pre-enrollment data to have its maximum effect. The usual first semester rate of nonpersistence at The University of Iowa is about 7%. Although this is not as high as the second semester nonpersistence rate, the second semester was assumed to be influenced more by factors other than the pre-enrollment data.

Specific objectives of this study were to :

(1) identify pre-enrollment characteristics that distinguish non-persisting students from students who persist into the second semester of their freshman year,

(2) reduce the large number of pre-enrollment variables to a small number of useful prediction variables, and

(3) use linear and nonlinear classification techniques to thoroughly exhaust the predictive power of the prediction variables.

PROCEDURES

Subjects

The Fall 1978 class of entering freshman at The University of Iowa served as the primary sample in this study. About 7% of the 1978 cohort dropped out before the start of the second semester of the 1978-1979 academic year. This rate is similar to those observed in past years. The 1978 cohort was grouped as follows :

-Group A: "Non-persisters"---the 7% of the freshmen who left the University by the second semester of their freshman year;

-Group B: "Persisters"-------those freshmen in the 1978 cohort who started the second semester of their freshman year.

Separate analyses were not done for males and females due to the small sample size of group A. Previous studies have indicated that there is no difference in persistence for males and females. The Fall 1979 class of entering freshman at The University of Iowa served as the cross validation group.

Data Sources

Three sources of information were used in this study to obtain predictor variables : ACT test scores, the ACT Student Profile, and high school transcript information from the University Registrar. The ACT tests measure academic skills in English, mathematics, natural science, and social science. A composite test score based on performance in these areas is also provided. The Student Profile Report provides information about the student's background, past accomplishments, special needs, and future educational/vocational objectives. Also provided in the Student Profile Report are special scores from an ACT Interest Inventory. Information used from the high school transcript includes grade point average, class rank, and the size of the graduating class. In most cases, the high school transcript information and all the ACT data were provided to The University of Iowa as a requirement for admission. Using data directly from the above sources, nine combined variables were created. These variables were simple sums of binary response items relating to specific areas of students' extracurricular activities in high school.

Analysis

A Pearson correlation matrix was produced using the pre-enrollment information from the 1978 cohort. Missing values were not replaced while developing the correlation

matrix. As a result, the number of cases used to calculate the correlations ranged from 46 to 2842, depending on the variables correlated. The correlation matrix formed the basis for the subsequent calculations.

A Principal Components analysis was performed on the correlation matrix to produce eigenvalues and to determine the portion of variance each factor contributed to the overall variance. Based on these results, two sets of factors were chosen. The first set consisted of the fewest number of factors that could account for a substantial proportion of the variance, arbitrarily determined to be 65%. Forty-three factors were retained using the above criterion. If a minimum eigenvalue of 1 been used as the criterion, forty-one factors would have been retained. The second set consisted of those factors that contributed at least 1.1% to the total variance. The criterion was chosen after reviewing the criterion produced using the Scree method of factor selection. The second set of factors was produced because forty-three factors was considered too large a number of factors to retain. Twenty-three factors were retained using this second approach.

Both sets of factors were rotated using varimax rotation and factor scores were calculated for each student. Two discriminant classification functions were calculated using the two different sets of factor scores. The classification variable was first semester persistence, coded as a

zero or a one. A test of homogeneity was made for each function to determine whether the pooled covariance matrices for the discriminant functions could be validly used.

Cross validation was performed on the 1979 cohort. The factor scoring patterns from each factor analysis were used to calculate the factor scores for the 1979 cohort. Both the 43- and 23-factor scores were obtained. The discriminant functions calculated for the 1978 cohort were then applied to the 1979 factor scored data.

Finally, in order to account for possible nonlinearity in the data, a near neighbor analysis was performed on the 43-factor data. Only part of the analysis was completed due to the high cost. The incomplete analysis yielded poorer results than the discriminant results on the calibration data.

Results

There were 2,850 entering freshman at The University of Iowa in the fall 1978 cohort. Although 132 pre-enrollment variables were theoretically available for each student, there were missing values. Eleven of the pre-enrollment variables had less than 2000 respondents. To increase the number of usable cases in both the 1978 and 1979 cohort, averages were substituted for missing values for these eleven variables. Thus 1,711 of these students had enough information to allow factor scoring.

The results of the first factor analysis used to produce eigenvalues and to determine portion of variances are given in Table 1. The sets of 43 and 23 factors were obtained from this factor analysis.

The rotated 43-factor set accounted for 65% of the overall variance. The calibration data applied to the discriminant function yielded the results shown in Table 2. The accuracy of the prediction was 95.73%, the actual rate of retention was 93.73%. The test of homogeneity for the within covariance matrices was rejected at the $p<.001$ level with a chi-square value of 1288 and 946 degrees of freedom. When the 2801 students of the 1979 cohort were factor scored, 1697 of the subjects remained for analysis. These subjects were then entered into the discriminant function. The results of this procedure are shown in Table 3. The discriminant function had an accuracy of only 90.51% when applied to the 1979 cohort.

The factor analysis retaining only 23 factors was also varimax rotated. The amount of overall variance accounted for was 48%. Table 4 shows the results of the calibration data applied to the discriminant function. Table 5 shows the cross validation data applied to the discriminant function. The accuracy of the calibration data was 92.57%. The accuracy of the cross validation data was 89.16%. The homogeneity of within covariance matrices was rejected at the $p<.001$ level as the chi-square value was 489 with 276 degrees of freedom.

DISCUSSION

The discriminant results of the two analyses were marginally better than chance for the calibration data. The cross validation results are worse than chance. This follows the pattern of previous retention studies at The University of Iowa. Variables that predict the calibration data better than chance fail to do so on cross validation data.

On the basis of the cross validation results, the difference between using 43 factors and using 23 factors was negligible. If any variables are to be used from the pre-enrollment data, the factor scores from the 23 factor analysis should be used. There were no indications from this or previous studies that a larger number of these pre-enrollment variables could account for more of the variance of persistence. Previous studies at The University of Iowa did no better than the results obtained here. The use of 23 factors rather than all 132 variables would considerably reduce the costs of analyses using the ACT profile, the ACT scores, and registrar's high school information. However, the 23 factors do not predict retention any better than an earlier study which used only the 3 variables : High school GPA, college preparatory curricullum in High school, and student confidence in future academic or vocational plans. This seems to indicate that even 23 factors are too many and that a smaller set of factors should be used. This result may also indicate that the pre-enrollment variables may not be able to predict persistence with acceptable accuracy.

In conclusion, it appears that pre-enrollment variables per se do not directly predict persistence. Therefore, current research at Iowa is based on information sources other than the pre-enrollment variables. The smallest set of factors that maintains the current level of prediction of persistence should be developed and used in retention studies to save on costs of storage and costs of analysis, but in conjunction with other information. It should be noted that it may be possible to affect the rate of persistence by intervention on the basis of pre-enrollment data, but that has not been investigated here.

TABLE 1

	1	2	3	4	5	6	7	8	9	10	11	12
EIGENVALUES	9.657069	6.154212	4.487497	3.932875	3.178580	2.822789	2.662298	2.532292	2.361168	2.259148	2.162618	2.078064
PORTION	0.073	0.047	0.034	0.030	0.024	0.021	0.020	0.019	0.018	0.017	0.016	0.016
CUM PORTION	0.073	0.120	0.154	0.184	0.208	0.229	0.249	0.268	0.286	0.303	0.320	0.336
	13	14	15	16	17	18	19	20	21	22	23	24
EIGENVALUES	1.973875	1.923205	1.873593	1.822397	1.687877	1.670758	1.641204	1.585975	1.559628	1.484623	1.444975	1.382393
PORTION	0.015	0.015	0.014	0.014	0.013	0.013	0.012	0.012	0.012	0.011	0.011	0.010
CUM PORTION	0.350	0.365	0.379	0.393	0.406	0.418	0.431	0.443	0.455	0.466	0.477	0.487
	25	26	27	28	29	30	31	32	33	34	35	36
EIGENVALUES	1.375753	1.338891	1.306040	1.265801	1.229773	1.197349	1.167180	1.161033	1.128310	1.106511	1.093261	1.079804
PORTION	0.010	0.010	0.010	0.010	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.008
CUM PORTION	0.498	0.508	0.518	0.527	0.537	0.546	0.555	0.563	0.572	0.580	0.589	0.597
	37	38	39	40	41	42	43	44	45	46	47	48
EIGENVALUES	1.067476	1.050165	1.026219	1.015509	1.004517	0.983476	0.961072	0.956004	0.940671	0.935508	0.908804	0.897376
PORTION	0.008	0.008	0.008	0.008	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
CUM PORTION	0.605	0.613	0.621	0.628	0.636	0.643	0.651	0.658	0.665	0.672	0.679	0.686
	49	50	51	52	53	54	55	56	57	58	59	60
EIGENVALUES	0.885786	0.864062	0.854960	0.847197	0.838763	0.831152	0.823051	0.816249	0.801357	0.789236	0.786252	0.772407
PORTION	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
CUM PORTION	0.693	0.699	0.706	0.712	0.718	0.725	0.731	0.737	0.743	0.749	0.755	0.761
	61	62	63	64	65	66	67	68	69	70	71	72
EIGENVALUES	0.763506	0.743453	0.740206	0.722920	0.716054	0.703301	0.684484	0.674379	0.672948	0.650701	0.642175	0.636387
PORTION	0.006	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
CUM PORTION	0.767	0.772	0.778	0.783	0.789	0.794	0.799	0.804	0.810	0.815	0.819	0.824
	73	74	75	76	77	78	79	80	81	82	83	84
EIGENVALUES	0.634022	0.616525	0.607513	0.602944	0.592797	0.586065	0.576585	0.575926	0.570797	0.561520	0.546048	0.540360
PORTION	0.005	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
CUM PORTION	0.829	0.834	0.838	0.843	0.847	0.852	0.856	0.861	0.865	0.869	0.873	0.877
	85	86	87	88	89	90	91	92	93	94	95	96
EIGENVALUES	0.533222	0.525122	0.524692	0.517245	0.507922	0.503044	0.496911	0.485986	0.473876	0.471344	0.459713	0.442852
PORTION	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003
CUM PORTION	0.881	0.885	0.889	0.893	0.897	0.901	0.905	0.908	0.912	0.915	0.919	0.922
	97	98	99	100	101	102	103	104	105	106	107	108
EIGENVALUES	0.437405	0.434558	0.424861	0.418165	0.407093	0.404088	0.394232	0.389277	0.376437	0.372399	0.367225	0.358703
PORTION	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
CUM PORTION	0.926	0.929	0.932	0.935	0.938	0.941	0.944	0.947	0.950	0.953	0.956	0.959
	109	110	111	112	113	114	115	116	117	118	119	120
EIGENVALUES	0.351742	0.346653	0.329002	0.325700	0.313952	0.312020	0.295871	0.289315	0.269469	0.263459	0.259875	0.249516
PORTION	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
CUM PORTION	0.961	0.964	0.966	0.969	0.971	0.974	0.976	0.978	0.980	0.982	0.984	0.986
	121	122	123	124	125	126	127	128	129	130	131	132
EIGENVALUES	0.244471	0.237526	0.232735	0.221181	0.219897	0.214365	0.204716	0.172802	0.136756	0.110566	0.004529	-0.137421
PORTION	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.000	-0.001
CUM PORTION	0.988	0.990	0.991	0.993	0.995	0.996	0.998	0.999	1.000	1.001	1.001	1.001

TABLE 2

43 FACTOR - DISCRIMINANT CALIBRATION RESULTS

		PREDICTED RETENTION		
		LEAVE	STAY	TOTAL
ACTUAL	LEAVE	74 63.25%	43 36.75%	117 100.0%
	STAY	30 1.88%	1564 98.12%	1594 100.00%
	TOTAL	104 6.08%	1607 93.92%	1711 100.0%

ACCURACY OF 95.73%.

TABLE 3

43 FACTOR - DISCRIMINANT CROSS VALIDATION RESULTS

		PREDICTED RETENTION		
		LEAVE	STAY	TOTAL
ACTUAL	LEAVE	7 7.29%	89 92.71%	96 100.0%
	STAY	71 4.43%	1530 95.57%	1601 100.00%
	TOTAL	78 4.60%	1619 95.40%	1697 100.0%

ACCURACY OF 90.57%.

TABLE 4

23 FACTOR - DISCRIMINANT CALIBRATION RESULTS

		PREDICTED RETENTION		
		LEAVE	STAY	TOTAL
ACTUAL	LEAVE	29 24.79%	88 75.21%	117 100.0%
	STAY	39 2.45%	1555 97.55%	1594 100.00%
	TOTAL	68 3.97%	1643 96.03%	1711 100.0%

ACCURACY OF 92.57%.

TABLE 5

23 FACTOR - DISCRIMINANT CROSS VALIDATION RESULTS

		PREDICTED RETENTION		
		LEAVE	STAY	TOTAL
ACTUAL	LEAVE	7 7.29%	89 92.71%	96 100.0%
	STAY	95 2.45%	1506 97.55%	1601 100.00%
	TOTAL	102 6.01%	1595 93.99%	1697 100.0%

ACCURACY OF 89.16%.

REFERENCES

Doolittle, Allen E. First-Year Retention of College Students. Report No. 81-102, April 1981, Evaluation and Examination Service, The University of Iowa.

Pantages, R. J. and Creedon, C. F. Studies of college attrition : 1950-1975. Review of Educational Research, 1978, 48, 49-101.

Summerskill, J. Dropouts from college. In N. Sanford (Ed.), The American College. New york : Wiley, 1962.